



A Parallel Hybrid Solver for Large Sparse Linear Systems in End-to-end Accelerator Structure Simulations

Lie-Quan Lee, Lixin Ge, Marc Kowalski, and Kwok Ko
Stanford Linear Accelerator Center

SciDAC Collaborators at LBNL
W. Gao, P. Husbands, X. Li, C. Yang, and E. Ng

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Overview

- Parallel Simulation Codes
 - Omega3P/S3P/T3P
- Linear Solver Framework
- Hybrid Linear Solver
 - Conjugate Gradient with Hierarchical Preconditioner
- Wakefield Computations
- Work in Progress

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Acknowledgments

- Work supported by DOE's HEP and ASCR Offices under the SciDAC Project
- All Simulations performed at NERSC's IBM/SP



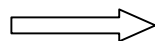
Overview

- Parallel Simulation Codes
 - Omega3P/S3P/T3P
- Linear Solver Framework
- Hybrid Linear Solver
 - Conjugate Gradient with Hierarchical Preconditioner
- Wakefield Computations
- Work in Progress

Omega3P

(Eigensolver for Finding Normal Modes)

Maxwell's Equations. In
Frequency Domain



Finite Element
Formulation

$$\nabla \times \left(\frac{1}{\mu} \nabla \times \mathbf{E} \right) - \frac{\omega^2}{c^2} \epsilon \mathbf{E} = 0$$

$$n \times \mathbf{E} = 0 \quad \text{on electric boundary}$$

$$n \times \left(\frac{1}{\mu} \nabla \times \mathbf{E} \right) = 0$$

on magnetic boundary

$$\mathbf{E} = \sum_i e_i \mathbf{N}_i$$

$$\mathbf{Kx} = \frac{\omega^2}{c^2} \mathbf{Mx}$$

$$\mathbf{K}_{i,j} = \int \frac{1}{\mu} (\nabla \times \mathbf{N}_i) \cdot (\nabla \times \mathbf{N}_j) d\Omega$$

$$\mathbf{M}_{i,j} = \int \epsilon \mathbf{N}_i \cdot \mathbf{N}_j d\Omega$$

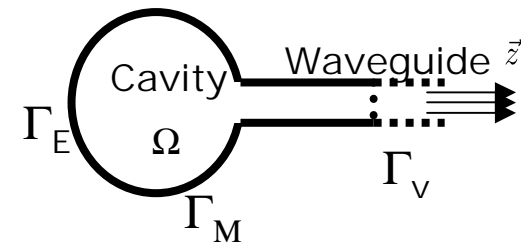
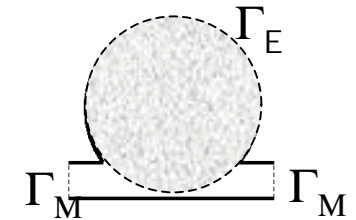
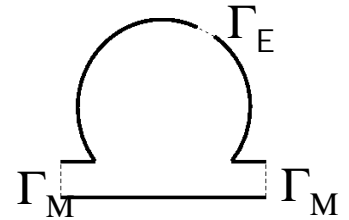
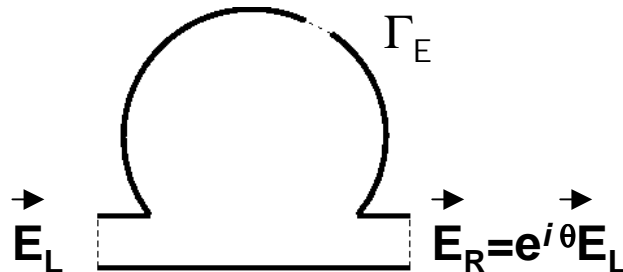
- Shift-Invert Lanczos (SIL):
 - need to solve shifted linear system $(\mathbf{K} - \sigma \mathbf{M})\mathbf{x} = \mathbf{b}$

Omega3P Matrices

(Application to Accelerator Cavities)

■ Matrix

- Real sparse symmetric (Closed cavity),
- Complex sparse symmetric (Lossy materials),
- Nonlinear (External Coupling),
- Complex sparse Hermitian (Periodical structures)



S3P

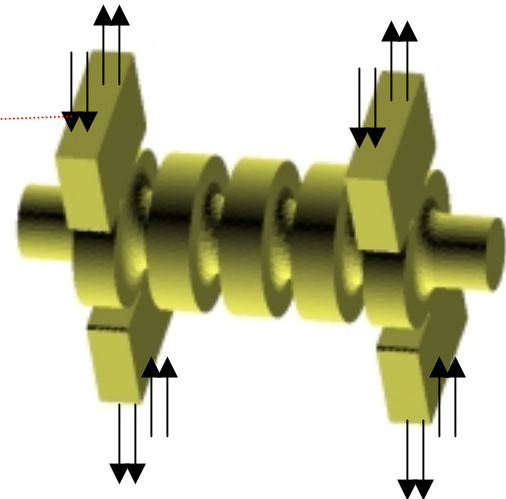
(Scattering Matrix Computations)

- Frequency-domain solver for finding the scattering matrix of traveling wave structures
- Matrix properties
 - Real symmetric (Lossless) or Complex symmetric (Lossy materials)

$$\int_{\Omega} \frac{1}{\mu_r} (\nabla \times \vec{E})^* \cdot (\nabla \times \vec{h}) d\nu - \frac{\omega^2}{c^2} \int_{\Omega} \vec{E}^* \cdot \vec{h} d\nu$$

$$= -i\omega\mu_0 \int_S (\vec{n} \times \vec{H}_{excit})^* \cdot \vec{h} ds$$

$$(\mathbf{K} - \frac{\omega^2}{c^2} \mathbf{M}) \cdot \mathbf{x} = \mathbf{b}$$



T3P

(Time Domain Simulation)

$$\mathbf{M} \frac{d^2 \mathbf{u}}{dt^2} + \mathbf{D} \frac{d\mathbf{u}}{dt} + \mathbf{K} \mathbf{u} = \mathbf{f}$$

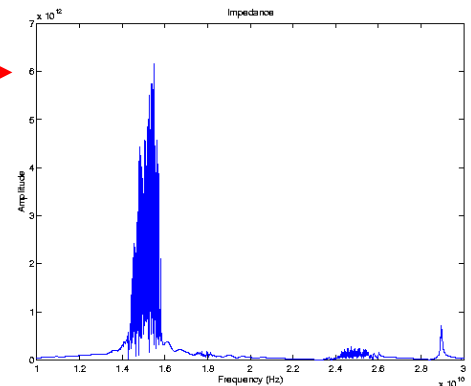
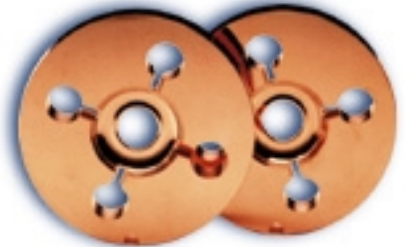
- Beam, dipole, waveguide port excitation
- Implicit time stepping scheme (Newmark-beta scheme)
- Need to solve a linear system each time step
- Long simulation time -- Multiple right hand sides:
100,000 to 1,000,000

Challenges in Accelerator Simulations

- High accuracy for complex geometry →
 - **0.01%** frequency accuracy to meet tolerance requirement

- Large-scale end-to-end simulation →
 - Discretization resulting in matrices of **10's to 100's million DOFs**

- Broadband response
 - Small beam excites a **dense, broad spectrum** consisting of hundreds to thousands of modes that are tightly clustered (0.5% separation).



A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Overview

- Parallel Simulation Codes
- **Linear Solver Framework**
- Hybrid Linear Solver
 - Conjugate Gradient with Hierarchical Preconditioner
- Wakefield Computations
- Work in Progress

Linear Solver Framework

Linear Solver Framework

Direct Solver Interface

Krylov-subspace Solvers

Preconditioners

Basic Linear Algebra Interface

■ Component-based Design

- Interface to SuperLU, WSMP, MUMPS
- CG, GMRES, QMR, SSOR, ILU, IC, ...
- Fortran BLAS, Boost μ BLAS, MTL, Blitz++

■ Extensible

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

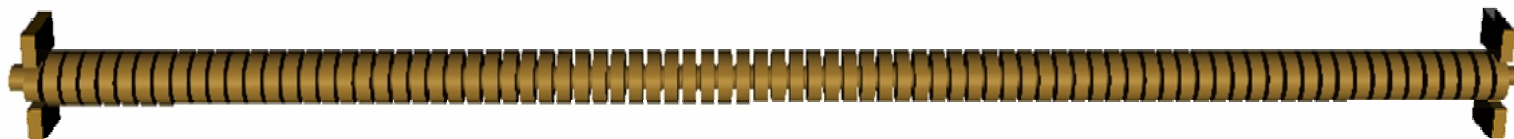
Overview

- Parallel Simulation Codes
- Linear Solver Framework
- Hybrid Linear Solver
 - Conjugate Gradient with Hierarchical Preconditioner
- Wakefield Computations
- Work in Progress

Motivation for Hybrid Linear Solver

H60VG3 is a 55-cell tapered structure considered to be the baseline design for the Next Linear Collider (NLC).

A coarse mesh is used for illustration purpose



- 1.3M DOFs / 4x16 CPUs on NERSC SP2 / 16 right hand sides

	CG+SSOR(1,4)	WSMP
Total Time	14582.1s	82.6s
Fact. Time	-	74.4s
Memory	2.9GB	19GB

Direct solvers are one to two orders of magnitude faster

Limitation of Direct Solvers

H60VG3 structure, linear element, $N=30M$, $nnz=484M$

S3P with WSMP

- 1024 CPUs, 487GB
 - Ordering time: 4248s
 - Numerical Factorization: 133s
 - Triangular solver (per RHS): 5.84 second

Omega3P with ESIL + WSMP

- 1024 CPUs, 738GB
 - Ordering time: 4143s
 - Numerical Factorization: 133s
 - Total: 5068s for 12 eigenvalues with 3 shifts

- Direct solvers require a large amount of memory per CPU
- What do we do for larger problems?

A decorative graphic on the left side of the slide, consisting of overlapping yellow, red, and blue squares with a black crosshair.

Hybrid Linear Solver

- Combine the benefits of both direct solvers and iterative solvers
 - Fast solving time (direct)
 - Less memory usage (iterative)
- Conjugate Gradient with Hierarchical Preconditioner (CGHP)
 - Implemented up to 6th order hierarchical finite element bases
 - Use solutions from direct solvers on the lower order system as preconditioner in CG for higher order system



Background

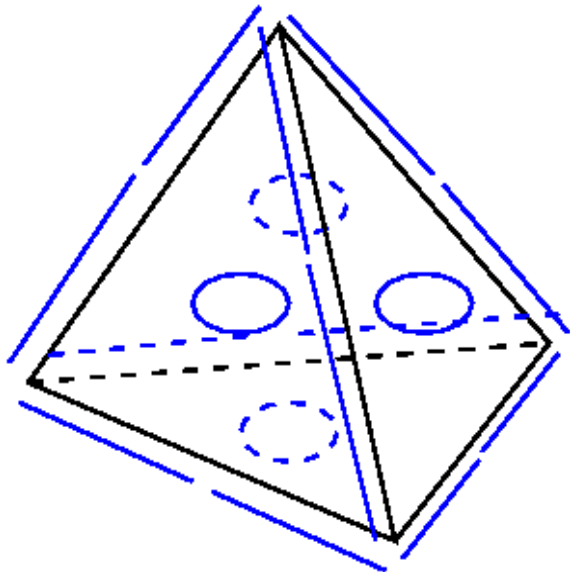
- Linear system $\mathbf{Ax} = \mathbf{b}$
- The convergence of iterative linear solvers such as Conjugate Gradient algorithm strongly depend on preconditioner used
- Preconditioner \mathbf{M}
 - \mathbf{M} is a good approximation of \mathbf{A}
 - It is easy to solve linear system $\mathbf{My} = \mathbf{z}$
- CGHP
 - The solution from the matrix assembled from FEM using lower order bases is projected back and used as solution of the preconditioner system
 - Direct solver is used for solving the lower-order system

Hierarchical Vector Bases

- **$p+1$** -order basis function set includes the **p** -order basis function set

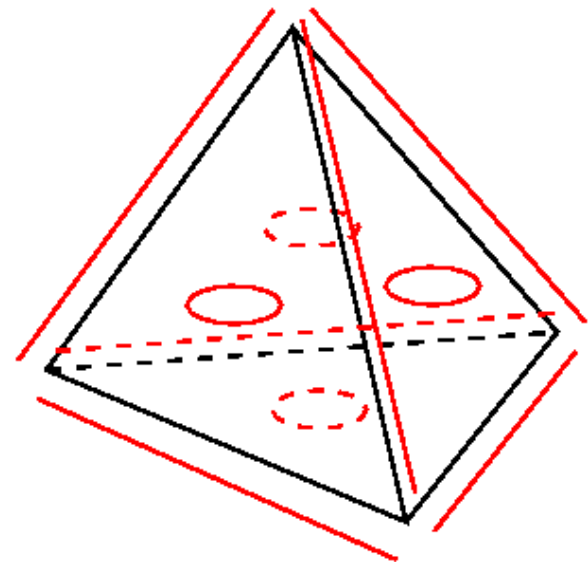
Black: linear

Blue: non-hierarchical quadratic



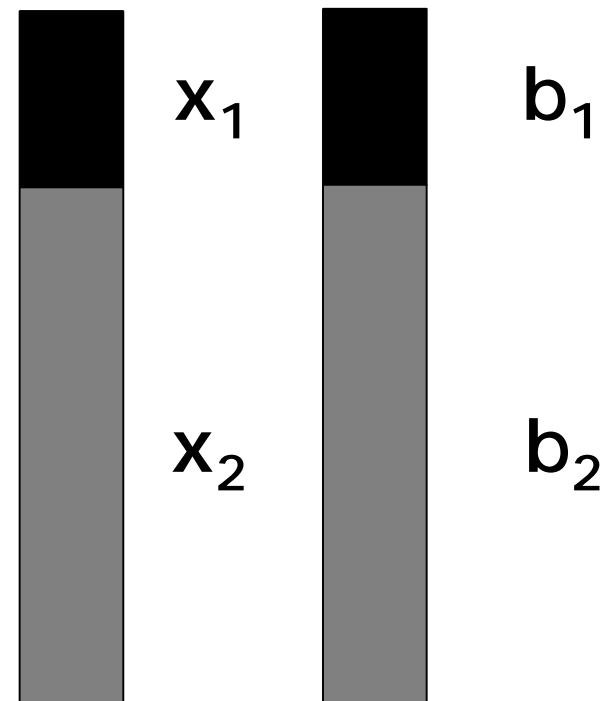
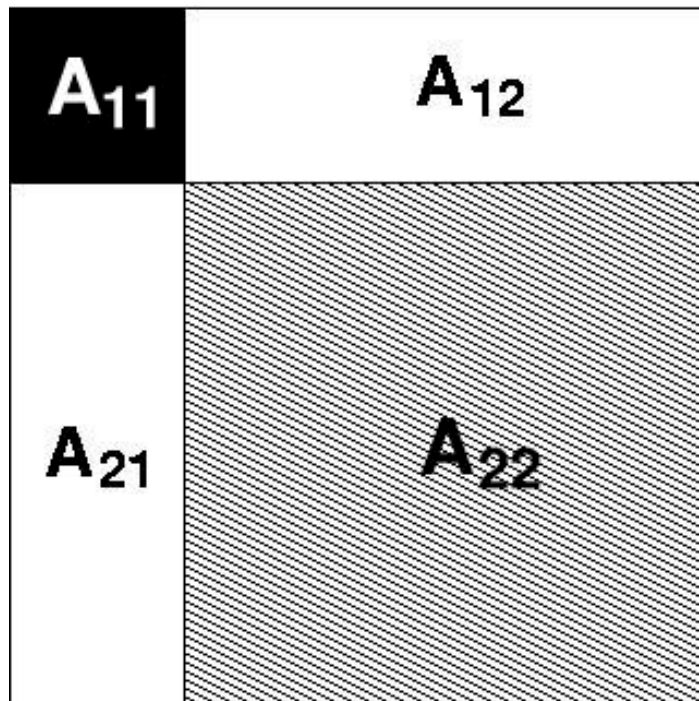
Black: linear

Red+black: hierarchical quadratic



Linear System and Numbering

- Solve $Ax=b$
- Number the p -order DOFs before $p+1$ -order DOFs





Block Jacobi

- $\mathbf{x}_1 = \mathbf{A}_{11}^{-1} \mathbf{b}_1$
- $\mathbf{x}_2 = \mathbf{A}_{22}^{-1} \mathbf{b}_2$


where the preconditioner is:

Direct factorization **\mathbf{A}_{11}**

\mathbf{A}_{22} SSOR, diagonal scaling...

Symmetric Block Gauss-Seidel

- $\mathbf{x}_1' = \mathbf{A}_{11}^{-1} \mathbf{b}_1$
- $\mathbf{x}_2 = \mathbf{C}_{22}^{-1} (\mathbf{b}_2 - \mathbf{A}_{21} \mathbf{x}_1')$
- $\mathbf{x}_1 = \mathbf{A}_{11}^{-1} (\mathbf{b}_1 - \mathbf{A}_{12} \mathbf{x}_2)$

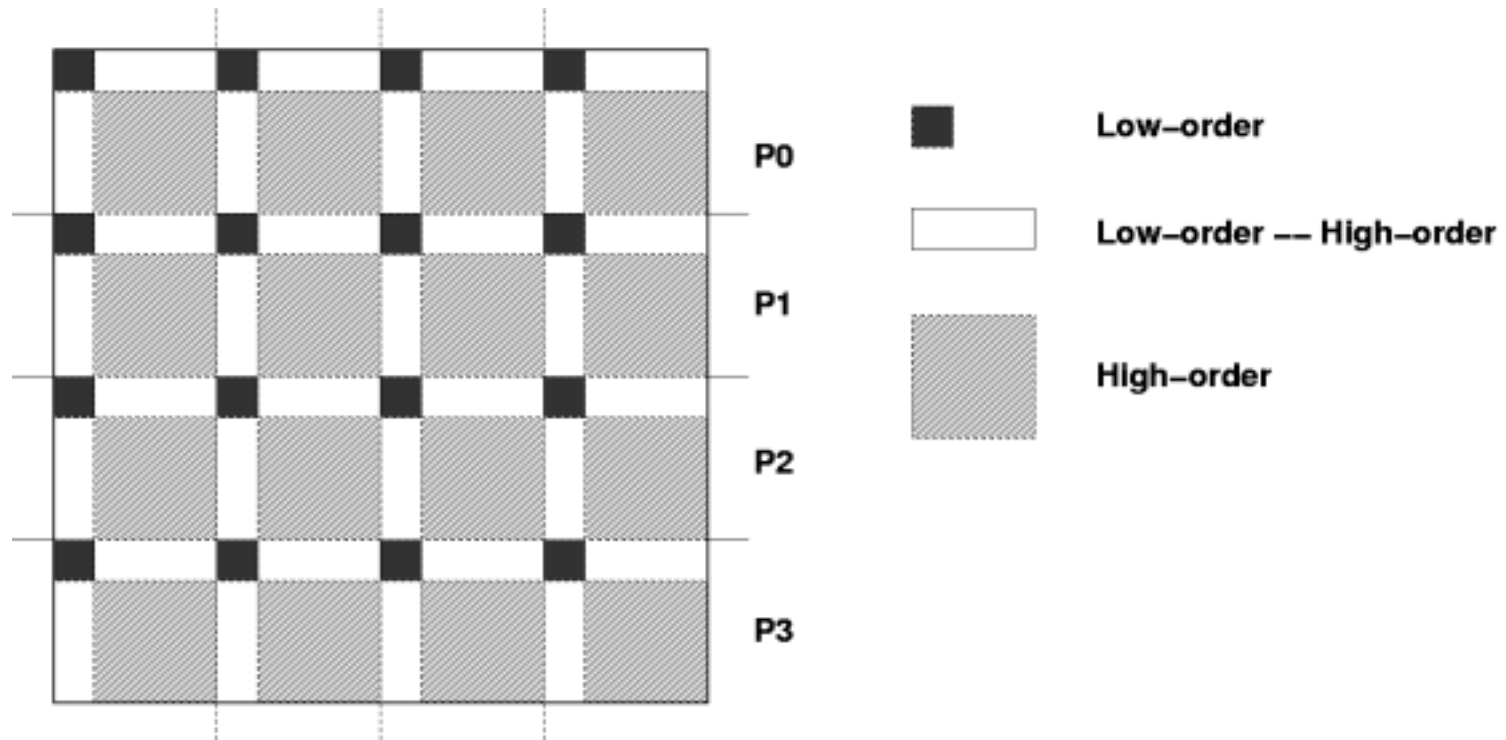

Direct factorization

$$\begin{bmatrix} \mathbf{A}_{11} & \\ & \mathbf{C}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \\ & \mathbf{C}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ & \mathbf{C}_{22} \end{bmatrix}$$

Where $\mathbf{C}_{22} = (\mathbf{D}_{22} + \omega \mathbf{L}_{22}) \mathbf{D}_{22}^{-1} (\mathbf{D}_{22} + \omega \mathbf{L}_{22}^T)$

Parallel Implementation of Hierarchical Preconditioner

- Number DOFs in two groups
- Avoid vector copying





CGHP Used in S3P

- H60VG3 S3P study
- 14 right hand sides
- 1024 CPUs on NERSC IBM SP2
- CGHP would be **faster** in solving 14 RHS than WSMP for the same problem size (assuming WSMP execution time scales linearly with problem size)

Solver	Prob. Size	Memory	Time (s)
WSMP	N=30M, nnz=484M	487GB	4462.9
CGHP	N=93M, nnz=4billion	836GB	6456.8



CGHP Used in Omega3P

- H60VG3 eigen-analysis, quadratic element, ESIL with CGHP
 - **n=93 million, nnz=4 billion**
 - 128CPUs for CG and ESIL iterations, 1024CPUs for WSMP
 - 2 shifts, 8 eigenvalues
 - Time: 1413s for ordering, 73s per factorization, 420min total
 - About 50 CG iterations per linear system (independent of mesh size)
 - Used **704GB** only (would be over 2.5TB using direct linear solver)

A decorative graphic on the left side of the slide, consisting of overlapping yellow, red, and blue squares with a black crosshair.

Summary for CGHP

- Memory efficient
 - Can solve bigger problems
- Performance is comparable to direct solvers
- Convergence is independent of mesh size
- Hierarchical high order bases needed

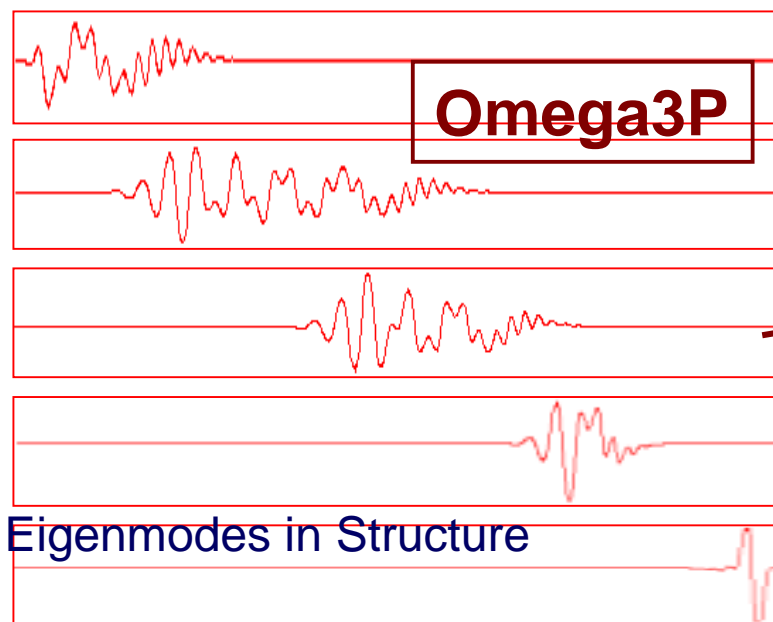
A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Overview

- Parallel Simulation Codes
- Linear Solver Framework
- Hybrid Linear Solver
 - Conjugate Gradient with Hierarchical Preconditioner
- Wakefield Computations
- Work in Progress

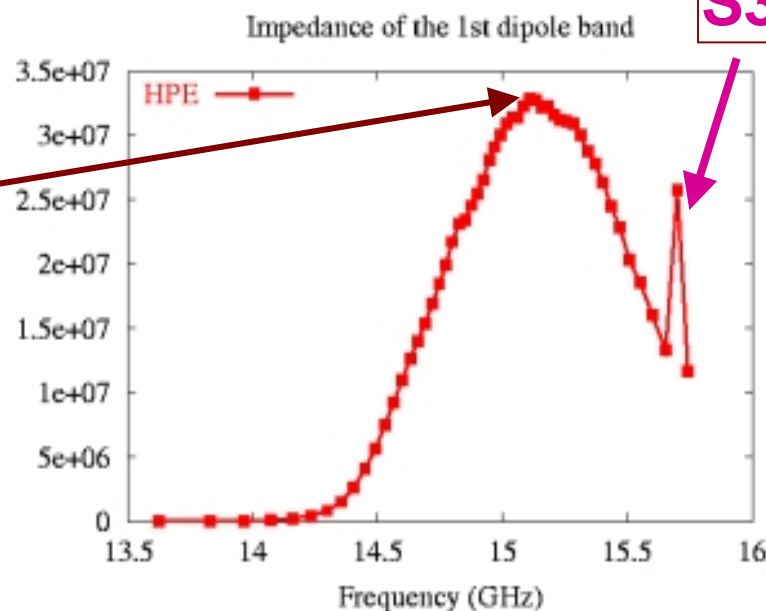
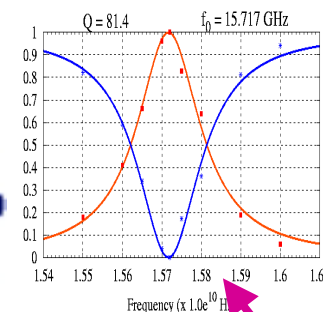
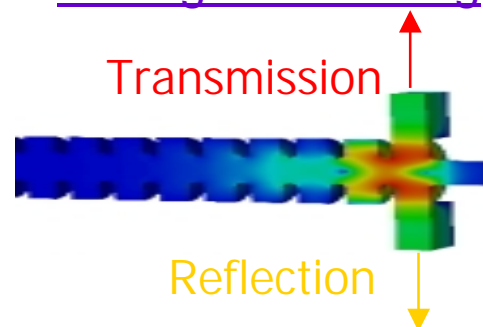
Wakefields via Mode Analysis

Omega3P – Accurate calculations of tightly-clustered eigenvalues (0.5% separation) in broad spectrum



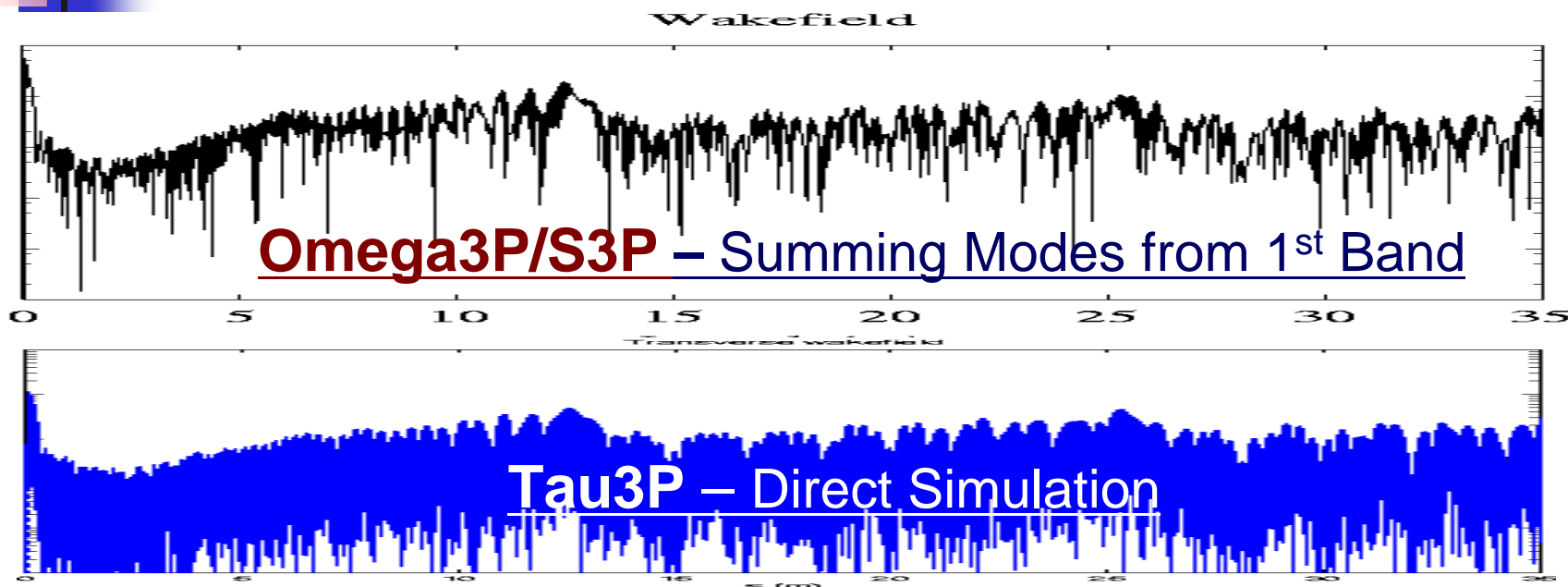
Eigenmodes in Structure

Waveguide Loading



S3P

Detuned Structure Wakefields



- First ever direct comparison between time and frequency domain calculations of wakefields in a realistic structure,
- Demonstrate that system scale simulation possible with parallel computing and valuable for accelerator R&D.

A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

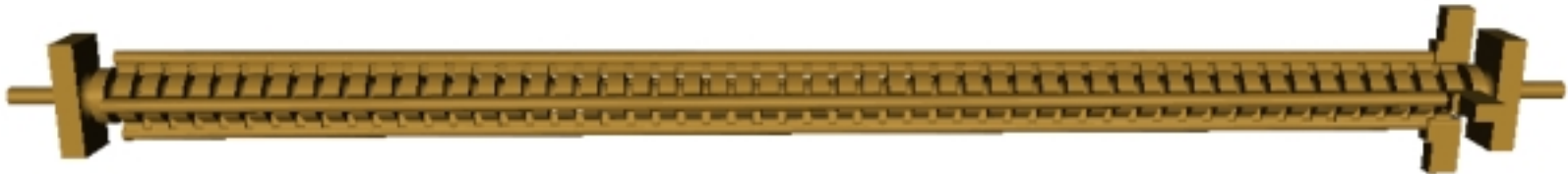
Overview

- Parallel Simulation Codes
- Linear Solver Framework
- Hybrid Linear Solver
 - Conjugate Gradient with Hierarchical Preconditioner
- Wakefield Computations
- Work in Progress



The Next Big Challenge

- Simulate H60VG3 with damping
(Damped Detuned Structure)
- Develop parallel *complex eigensolver*
- Estimated matrix sizes:
 - $N \geq 200$ millions
 - $NNZ \geq 8$ billions



A decorative graphic consisting of overlapping yellow, red, and blue squares with a black crosshair.

Work in Progress

(Involving SLAC, LBNL and Stanford as part of DOE's Accelerator Simulation SciDAC Project)

- Study the impact to linear solvers, eigensolvers by removing null space
- Improve the scalability of parallel solutions of sparse triangular linear systems
- Develop new direct, iterative or hybrid solvers for large sparse symmetric indefinite linear systems that have multiple right-hand sides